DOCUMENT RESUME

ED 059 630 24 EM 009 704

AUTHOR	Collet, LeVerne S.; And Others
TITLE	FEHR-PRACTICUM: A Computerized Game to Simulate Experience in Educational Research and Evaluation. Final Report.
INSTITUTION	Michigan Univ., Ann Arbor.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C.
BUREAU NO	BR-0-9047
PUB DATE	Aug 71
CONTRACT	OEC-0-70-4773(520)
NOTE	59p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Computer Assisted Instruction; Decision Making; Educational Development; *Educational Research; *Formative Evaluation; Problem Solving; *Simulation

ABSTRACT

This project represents the first year of a proposed two-year program to develop FEHR-PRACTICUM (Formative Evaluation and Heuristic Research), a computerized game which simulates experience in a research evaluation assistantship or practicum. In the game teams of from two to five players are given the task of finding the "best" among several specified educational alternatives. To accomplish this task, the teams perform "experiments" on a hypothetical school system simulated by the computer. When the team has decided on the best solution, they enter their decision into the computer, which then operates the entire hypothetical system for a year, and generates a "payoff matrix" comprised of the year-end mean values for a large set of dependent variables and the total additional cost imposed on the system by operating under that decision for one year. These values are than used to declare a winner for the game and to illustrate the relative efficacies of the analytic strategies used by the various teams. (Author:JY)

PA 24

ED 059630

FINAL REPORT

Project No. 0-9047

Contract No. OEC-0-70-4773 (520)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

FEHR-PRACTICUM:

A Computerized Game to Simulate Experience in Educational Research and Evaluation

LeVerne S. Collet
Director & Principal Investigator

With

David D. Starks
D. Lynne Feagans
Robert Fine
Robert Parnes
Sr. Eileen Rice

The University of Michigan
Ann Arbor, Michigan

August, 1971

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

EM009704

ABSTRACT

This project represents the first year of a proposed two-year program to develop FEHR-PRACTICUM, a computerized game which simulates experience in a research/evaluation assistanship or practicum. In the game, teams (2-5) are given the task of finding the "best" among several specified educational alternatives. To accomplish this task, the teams perform "experiments" on an hypothetical school system simulated by the computer. When the team has decided on the best solution, they enter their decision into the computer, which then operates the entire hypothetical system for a year, and generates a "payoff matrix" comprised of the year-end mean values for a large set of dependent variables and the total additional cost imposed on the system by operating under that decision for one year. These values are then used to declare a winner for the game and to illustrate the relative efficacies of the analytic strategies used by the various teams.

The specific product of this one-year project was an operational game model, together with validating evidence. The development and validation of ten problems to be used in the game is the object of a sequel project which the authors hope to undertake during 1971-72.

FINAL REPORT

Project No. 0-9047

Contract No. OEC-0-70-4773 (520)

FEHR-PRACTICUM:

A Computerized Game to Simulate Experience in Educational Research and Evaluation

LeVerne S. Collet
Director & Principal Investigator

With

David D. Starks
D. Lynne Feagans
Robert Fine
Robert Parnes
Sr. Eileen Rice

The University of Michigan
Ann Arbor, Michigan

August, 1971

TABLE OF CONTENTS

INTRODUCTION 1

Meaning of FEHR 1
Unique Aspects of This Report 2
Experiences Provided by the Game 3
Tasks Remaining 7
Need for the Game 8
Populations Served 9

SECTION I. A GENERAL DESCRIPTION OF THE FEHR-PRACTICUM GAME 11

Overview of the Game 11

Types of Problems 12
Components of the Game 13

Information Bank 13
Data Generator 13
Message Generator 15
IST Units 16

Operating Staff 16

Game Manager 16
Research Consultant 17

Learning the Game 17
Playing the Game 18

SECTION II. TECHNICAL DETAILS 24

The Message Generator 24

Description 24
Operation of the Message Generator 26
Status 27
Direction for Evaluation Year 27

FEHR-PRACTICUM Data Generator 29
I Database 29
Algorithm 30
Status of Database 32
Validation of Database 32

Rationale for the Generation of Rawscore Matrices 35

II Surveys and File Searches 37

Status of Survey 38
Validation 38

III Treatments 38

Status 42
Validation 42

IV Decision 43

Status 44
Validation 44

Portability of the Computer Program 45

Development of Problems 46

I. Selection of Problem Areas: Criteria 46
II. Problem Description Packages 46
III. Developmental Process 47
IV. Data for the Information Bank 47
V. Internal Representation in the Computer 48

Trial Operation 48

Summary 50

Training Educational Decision-Makers 50
Studying Problem Solving 51
Behavior Modification 51

References 52

LIST OF TABLES

TABLE 1. EFFECTS OF CHANGING NOMINAL WEIGHTS 34

LIST OF FIGURES

FIGURE I. SUMMARY CHART OF FEHR-PRACTICUM GAME 19

INTRODUCTION

There is a consensus among educators that practical on-the-job experience is an important component of programs to train educational researchers and evaluators. Empirical support for this argument was given by Buswell and McConnell (1966), who reported that involvement in research during training was the best single predictor of a student's future research contribution. The typical vehicles for providing practical experience in the past have been the research assistantship or practicum. However, these program components are generally too far from the classroom (in both time and distance) to provide efficient reinforcement for the skills and concepts learned in class. In fact, students commonly complain that there is little direct transfer from formal research design and statistics courses to the practical research experience.

The purpose of this project was to develop a computerized game which would simulate a practicum experience for the training of educational researchers and/or evaluators. Used in a research training program, the game would permit instructors to provide practice in the practical application of the skills learned in class. In addition, the game should prove to be a valuable tool for giving in-service evaluation training to personnel involved in funded research, curriculum development, and the like.

Meaning of FEHR

The proposed game is specifically intended to develop skill in the techniques of formative evaluation and heuristic research (FEHR)

Formative evaluation (after Scriven, 1963) refers to a continuing on-the-job assessment which performs the functions of feedback, diagnosis and guidance. Heuristic research, on the other hand, is meant to suggest an emphasis on practical solutions and performance criteria, on short-term field studies conducted to facilitate an educational decision. Consequently, the FEHR-PRACTICUM game features problems which require the players to plan, design and execute a series of field studies in the context of an operating (hypothetical) educational system.

Unique Aspects of This Report

This final report is unique in that it is actually a report of the progress made during the first year of a two-year project. The original FEHR-PRACTICUM proposal, submitted in June 1970, outlined a two-year program to develop a computerized game which would simulate a practicum experience for the training of educational researchers and/or evaluators. However, due to budgetary considerations at the U.S. Office of Education, it was decided to fund the project as two separate one-year projects, with the contract for the second year contingent upon successful completion of the first year. Since a new one-year proposal was not requested, it was necessary to derive the specific nature of the tasks to which we were committed from the original proposal. The tentative schedule in the original proposal (p. 21) listed the work to be completed during the first developmental year of the contract. In general, we were committed to develop each of the proposed physical components of the game, and to demonstrate that the game as a whole, in fact, works. We were not committed to the development of the internal

models (sets of parameter cards) which define the interrelationships among treatments, predictors and criteria for each of the ten problems during the first year.

We are pleased to report that all of the physical components are now fully operational. In addition, we have fully developed one problem (AUTOMATH) to demonstrate the flexibility of the model. The game as it presently exists is considerably more sophisticated than the model described in the original proposal. Specifically, the capacity of the game was increased in two ways. First, the computer program which generates the research data was modified to enable it to produce unique learning curves for each of the several thousand individual learners which comprise the population of subjects for a particular simulated research project. Thus it is now possible to simulate longitudinal studies or individual case studies. Secondly, a message generator was added to the game in order to simulate environmental events which may affect a research project. For example, the message generator might announce a one-month teacher strike. Such an event would normally involve additional expenses due to the delay. Frequently the players would be forced to completely revise their research plans because of the additional expenses and enforced changes in the time schedule.

Experiences Provided by the Game

FEHR-PRACTICUM is intended to provide a wide range of practical experience in educational research and evaluation without the expenses and time commitment involved in real research. Given the goal of simulating the research experience, common sense would dictate that the closer the simulation is to reality, the more

valuable will be its contribution to practice. Consequently the game is designed to provide many of the complex interactions (and frustrations) which are attendent on conducting research in schools as opposed to learning laboratories. The overall goal of the game is to provide players with experience in gathering and analyzing empirical data in order to arrive at a practical educational decision, and to provide players with feedback respecting the adequacy of their decisions. The game <u>as it presently exists</u> has the capacity to provide players with practical experience in each of the following areas:

1. Stating a problem in operational terms. The game provides considerable practice in this area since the computer requires all decisions classifications to be made in terms of the values of particular variables.
2. Preparing a budget and working within its constraints. Since players are given a finite research grant and must pay for each bit of information they collect, the game demands careful planning of expenditures.
3. Developing and following a sampling plan. The average FEHR-PRACTICUM problem contains literally thousands of potential research subjects, each with a wide variety of individual characteristics (sex, intelligence, socio-economic status, etc.). Almost any sampling plan which can be used in real research can be duplicated in the game--including plans which are invalid because of some type of selection bias.
4. Selecting dependent and independent variables which are relevant to a given problem and choosing the instruments (tests) which will be used to measure them. Although the players cannot devise their own tests, they may choose from a large pool of tests which are made

available in the game. To help them in assessing the utility of the various tests, players have access to test descriptions of the sort provided by Buros (1965). The game could provide several hundred separate test scores on each subject in the theoretical population. There is no limitation on the number of subjects, but the largest population used to date is 24,000.

5. Using survey techniques to identify the important dependent and independent variables in a given educational problem, and to clarify the relationships among variables. In the game, surveys are frequently required to determine the extent of a problem. For example, we may need to know how many students are retarded readers.

6. Designing research plans which isolate the effects of specific given educational treatments and treatment combinations. The game allows players to collect data according to almost any research design which can be used in a real-life situation--including biased or invalid designs. The list of possible designs includes N-way analysis of variance, latin squares, incomplete blocks, longitudinal studies (panel data), and case studies based on variable scores rather than verbal descriptions. Because of the capacity to produce longitudinal data, it is possible to simulate formative evaluation studies involving <u>sequences</u> of treatments and repeated observation periods.

7. Analyzing data collected from complex designs. The capacity to provide such analytic experience is ensured by the complex designs mentioned in (6) above. In addition, the game has a number of built-in biases which encourage players to use designs involving multiple criteria (dependent variables). Some training in using <u>patterns</u> of results (rather than the score on a single dependent variable) to assess an educational treatment is provided.

8. Modifying research plans to accommodate unforeseen events in the environment. For example, a teacher strike could cause an expensive delay in a project. Such events can be simulated with the present game.

9. Selecting consultants and preparing plans to optimize their effectiveness. The game provides an opportunity for players to explore their own limitations, and to find the conditions under which a consultant is "worth the money."

10. Relating the results of an evaluation to the time at which the evaluation is taken. The game permits program evaluations to be made for terms as short as one week or as long as 10-20 years.

11. Working with educational problems in a wide variety of content areas and at numerous educational levels. The topics of the ten problems provided with the present game run from the traditional subjects (e.g., mathematics and reading) to the specialized difficulties of handicapped children, and the levels represented include both pre-school children and college students.

12. Using self-help devices to improve their (the players') skills. Two aspects of the game are of interest here. First the game provides feedback on the quality of each educational decision and the validity of the information (experimental results) upon which the decision was based. Second, there are built in provisions for in-service training of individuals who are playing the games. The game contains a set of programmed materials to provide refresher courses for such topics as "choosing a valid research design," and "writing formation statements for canned statistical programs." In addition, players may "hire" a consultant to help them solve a specific problem--a great deal of incidental learning is likely to accrue from such player-consultant interaction.

13. Relating the quality of their decisions (as revealed by the feedback mentioned above) to their experimental procedures. The game simulates such experimental phenomena as:

    a. The effects of various levels of test reliability.
    b. The effects of biased sampling.
    c. The effects of moderator variables.
    d. The interactive effects of teachers and methods.
    e. The carryover effects from one treatment to another.
    f. The effects of the inflation of costs during an experiment.
    g. The interactive effects of testing (e.g., pre-tests) and treatments.

Tasks Remaining

Having outlined the capacities of the FEHR-PRACTICUM game, it behooves us to enumerate the tasks which must be completed before it is a usable product. In general, these include all the validation and field-testing tasks assigned to the second year of the project in the original proposal. A brief summary of the remaining tasks is outlined below.

1. The computer program must be revised to provide neatly-formatted printout which can be used by players directly. At present the game manager must supply a key to help the player identify his requested variables.
2. The parameter cards which define the internal state for each specific problem must be prepared and tested for verisimilitude.
3. The computer program must be tested for system-to-system portability. This will probably involve the preparation of several versions of the present program.
4. All prototype materials must be revised on the basis of the findings of the July pilot study. Final copies will not be prepared until we have extensive experience with the revised prototypes.

5. The effect of playing the game on research competence must be assessed empirically.
6. The preparation of a final report. It is planned that the report will include a comprehensive game manual.

Need for the Game

There is a consensus among educators that practical on-the-job experience is an important component of programs to train educational researchers and evaluators. Empirical support for this argument was given by Buswell and McConnell (1966), who reported that involvement in research during training was the best single predictor of a student's future research contribution.

The typical vehicles for providing practical experience in the past have been the research assistantship or practicum. However, these program components are generally too far from the classroom (in both time and distance) to provide efficient reinforcement for the skills and concepts learned in class. In fact, students commonly complain that there is little direct transfer from formal research design and statistics courses to the practical research experience. In addition, the amount and variety of practical experience which an institution can offer is severely limited by legal-ethical restrictions, the disproportionate amount of time required for data collection, and the difficulty of securing data which will provide practice in recognizing and interpreting important phenomena such as teacher-method interactions.

It is clear that there is a need for experiential training components which are easily accessible and more amenable to instructor management than the traditional research assistantship

or practicum. The data management aspects of the practicum have been effectively simulated by computer programs such as SIMEX (see Collet, 1969). Such programs allow one to generate, in a few minutes, data which closely resembles that obtained from several months of arduous data collection in an actual experiment. Furthermore, the computer can simulate any desired main effect or interaction, or display the effects of changes in criterion reliability and other such phenomena at will. The FEHR-PRACTICUM project seeks to extend the ideas developed in SIMEX to the simulation of the entire practicum experience.

The FEHR-PRACTICUM game is intended to be used as supplement to courses in research and evaluation methodology. It is not intended as a substitute for real research experience, but rather as an intermediary between the classroom and the work experience. The major advantage of a simulated practicum is its capacity to expose students to a wide variety of "practical" experiences in a short time. Such variety is unlikely to be obtained through assistantships or practicums in the time available to the trainee.

Populations Served

The game proposed here was intended to be used in one of two ways: (1) in training researchers or evaluators at the M.A. or Ph.D. level; and (2) giving in-service evaluation training to personnel involved in funded research, curriculum development, research units attached to educational systems, and the like. However, a major advantage of the game model proposed is that it is not content dependent. For example, the game as described could be

used to train teachers, psychologists, and others to use the patterns of scores on a battery of tests for individual assessment. With minor modifications, the game could be adapted to train teachers in various pedagogical strategies.

There are two possible restrictions on the use of FEHR-PRACTICUM. First, since the game is computer-simulated, the users have access to a computer which will accept the FORTRAN IV language and has at least a 32K memory unit. However, this is not considered a major disadvantage, since computers of that character are now widely available. Even outlying areas can use a portable terminal to contact a computer by telephone. A second potential restriction lies in the shortage of trained personnel to operate the system. The proposed solution is to program the computer to tell the operator exactly what to do. For systems not supporting interactive mode (remote terminals), the prompting units can be printed as a programmed learning package. In this way, the game could be played using computers which accept only IBM cards (batch process). The experience of the author with SIMEX, the forerunner of the simulation program in FEHR-PRACTICUM, suggests that students with no previous computer experience can be taught to generate simulated data in about thirty minutes.

Section I. A GENERAL DESCRIPTION OF THE FEHR-PRACTICUM GAME

FEHR-PRACTICUM is a game which provides experiences similar to those which you would have as a member of a team doing program research and evaluation for an educational system such as the public schools or a community college. The name FEHR was chosen because the game emphasizes formative evaluation and heuristic research; that is, evaluation conducted to provide feedback and guidance during program development, and research projects intended to provide information to facilitate specific decisions. Consequently, the FEHR-PRACTICUM experience features the planning and execution of field studies conducted in a hypothetical educational system. At the beginning of the first session, players are asked to participate in a game orientation program which will give them direct experience with each of the components described here. This is done by involving each player in a step-by-step solution for a sample problem. Once they have completed the orientation program, players will be thoroughly familiar with the game, and ready to tackle their first FEHR-PRACTICUM problem. A prototype copy of the orientation materials appears in the last section of this report.

OVERVIEW OF THE GAME

In the FEHR-PRACTICUM game, players (or teams of players) are given a specific educational problem which requires a decision to be made within a specified time interval. The players (teams) must then conduct simulated research projects in order to collect

information on which to base their decision. As in real life, research costs money. At the beginning of the game, each player (team) is given a research "grant" (in FEHR-PRACTICUM bank notes). Players must pay for each bit of information they collect during their research. Thus, one of their major tasks is to plan their research so as to ensure that they obtain sufficient information to permit a valid decision to be made without exceeding their grant funds.

Types of Problems

At the beginning of the game each player (or team) is given a Problem Description booklet. The booklet specifies an educational problem, outlines the alternative educational treatments (programs) which may be used in its "solution," and provides a good deal of background information about the problem setting or environment. They are required to make a decision respecting the given treatments. In general, this task will take one of the following forms:

(1) Given a particular educational program and the objectives it is intended to accomplish, determine the extent to which the program is meeting its objectives, and decide whether the program be continued.

(2) Given a specific set of objectives and several alternative treatments (programs), determine which treatment or treatment combination best meets the given objectives.

(3) Given a particular objective defined in terms of observable learner behavior and a large set of treatments (e.g., type of reinforcer, frequency of reinforcer, showing a movie, and so on), determine a program

(sequence of treatments or treatment combinations) which will regularly and reliably produce the desired behavior.

Components of the Game

Physically, the FEHR-PRACTICUM game consists of four interacting components: an information bank, a data generator, a message generator, and an in-service training (IST) unit. The function of each of these components is described below.

Information Bank. The information bank is actually a cross-referenced file. Historical information and statistical data that would normally be available to a researcher working in an educational system are stored on 5x8 index cards. The cards are filed by title (subject content) in the usual manner, and then cross-referenced by key words and phrases so as to facilitate retrieval. The source of information, specified on each card, may be an office file or an officer of the system. As in real life, information obtained from an individual (e.g., the superintendent of schools) is both cheaper and less reliable than information obtained from a file.

Information stored in the bank may be purchased by the players at any point in the game. The player simply indicates the source he wishes to address and the question he wishes to ask. The details of this transaction are explained to players during an orientation session.

Data Generator. The data generator is a computer program which enables players to make an empirical assessment of the effects of each of the treatments (programs) involved in a given

problem. Conceptually, the program contains a set of research "subjects" (e.g., students) who are predisposed to react to each of a large set of treatments in a particular way. The players can select which treatment to use with each subject and can control the duration (time) of the treatment.

The simulation program permits players to identify the "students" in any subgroups within the system. For example, if one asked for all the girls in grade 7 at school #1, the computer would print out the names (ID numbers) of all the students in the hypothetical system who satisfied these selection criteria. The simulated research which can be conducted in FEHR-PRACTICUM differs from real research primarily in the means by which the researcher (player) may communicate with his research subjects. In real life, the researcher can gather information about his subjects by administering tests and questionnaires, observing the subjects' actions under controlled conditions, or even by engaging individual subjects in conversation. In the FEHR-PRACTICUM game, the researcher (player) can communicate with his subjects only through their scores on the "tests" which he administers. In addition, each "test" to be used must be identified in advance: researchers may not develop their own tests. However, this is not so restrictive as it may sound, since a large set of tests (usually from 20 to 50) are available for each problem. The researcher (player) simply chooses those tests which "measure" the variables which are relevant to his particular problem. Some of the available tests will measure dependent variables; that is, variables which are changed over time by an educational treatment (or treatment

combination). Others will measure independent variables such as sex. These are not themselves affected by the treatments, but they may modify a treatment's effect on a dependent variable. For example, a particular treatment may be much more effective with boys than with girls. However, many of the available tests may be irrelevant to the problem a player is asked to solve. In addition, there is a charge for each student score which is printed out. Consequently, the first step a player must take in designing an experiment is to decide which variable scores are necessary to his purpose.

Once players have identified the groups (sets of subject ID's) of interest in their problem, an experiment can be run by specifying different treatments for each group and "measuring" the change in the relevant variables. If their "tests" are carefully chosen and the treatments administered according to a valid experimental design, they will be able to infer the effects of each treatment and/or treatment combination from a statistical analysis of the test scores. Note that pre- and post-test designs may be simulated by administering a treatment to the same group with two different times. This is done by setting the pre-tests at zero and the time for the post-test at the length time required for the experimental treatment.

Message Generator. It frequently happens that a research project is radically changed by external events which the experimenter cannot anticipate or control. For example, a teacher strike which interrupts an experiment may change pupil attitudes as well as introducing costly delays. Such "acts of God" are

introduced into the FEHR-PRACTICUM game by the message generator. At various times during the game, players may be given a message by the game manager. Some of these will be relatively unimportant and require no action on their part. Others, however, may require them to make adjustments in their research plan. For example, a message that a player's research budget had been cut might necessitate the use of smaller samples. Such messages are intended to provide players with experience in dealing with the unexpected.

IST Units. The IST (In-Service Training) units are sets of semi-programmed materials designed to upgrade the skills of the participants within the game itself. Although it should be possible to construct training modules so comprehensive that the game becomes a self-sufficient research training device, that is not the purpose of the present project. Rather, we specifically restricted ourselves to the development of a few exemplary units which concentrate on the specific skills needed to play the game. It should be noted, however, that the training modules could easily be supplemented by live teaching or published materials consonant with the FEHR-PRACTICUM theme. A particularly good example of the latter is Harshbarger's (1971) decision-map approach to statistics.

Operating Staff

The FEHR-PRACTICUM game requires the services of a game manager and at least one research consultant. The function of each of these staff members is explained below.

The game manager is the general supervisor of the game. His major task is to act as liaison between the players (teams) and the various game components: all requests to the information bank, the IST Unit, or the data

generator pass through his hands. The details of these transactions are explained to players during the orientation session.

The research consultant in the FEHR-PRACTICUM game serves the same function that he would in real-life research. Whenever a player (team) is uncertain about any aspect of their research methodology, they may hire a consultant to help them. If a number of teams are competing in a particular game, several research consultants may be provided. The game manager will announce at the beginning of each game the number of consultants available. In addition, he will provide the players (teams) with a vita on each consultant to help them decide which person to hire for any one task. Players may hire a consultant at any point during the game, providing one is available--at any one point in time it is possible for all the consultants to be engaged by other players (teams).

Learning the Game

If the competing teams are new to the FEHR-PRACTICUM game, it is necessary to train the players to use the various game components efficiently. Two sets of materials have been prepared for this purpose. The first of these is an illustrated booklet entitled Players' Introduction to FEHR-PRACTICUM. The first part of the booklet contains a brief description of the game and how to play it. The second half describes Fair City, the community

which provides the setting for each of the ten problems. The booklet also includes an extensive appendix containing census-type information about the various components of the city. For example, the size and location of each school in the city would be listed, along with the socio-economic characteristics of the pupils. The scores produced by the data generator will be consistent with this information in every respect.

The second set of materials is labelled the Programmed Orientation Booklet. These materials involve each player in a step-by-step "solution" to a sample problem. The sample game has been carefully structured to provide direct experience with each of the game elements. The general strategy for teaching each operational skill is: (1) present an example, (2) ask each player to perform the operation, and (3) provide feedback on the adequacy of the operation. Steps (2) and (3) are repeated several times for each skill to be taught.

Playing the Game

The details of playing a game of FEHR-PRACTICUM are, of course, explained fully during the orientation session. The intent of this section is to provide a general outline of a player's tasks during the game. The amount of time needed to play a FEHR-PRACTICUM game will vary depending on the research sophistication of the players and the length of time required by the particular computer system being used to generate and/or analyze the data. Consequently, the game would usually be played in four sessions of roughly two hours each, with the computer simulation runs and all

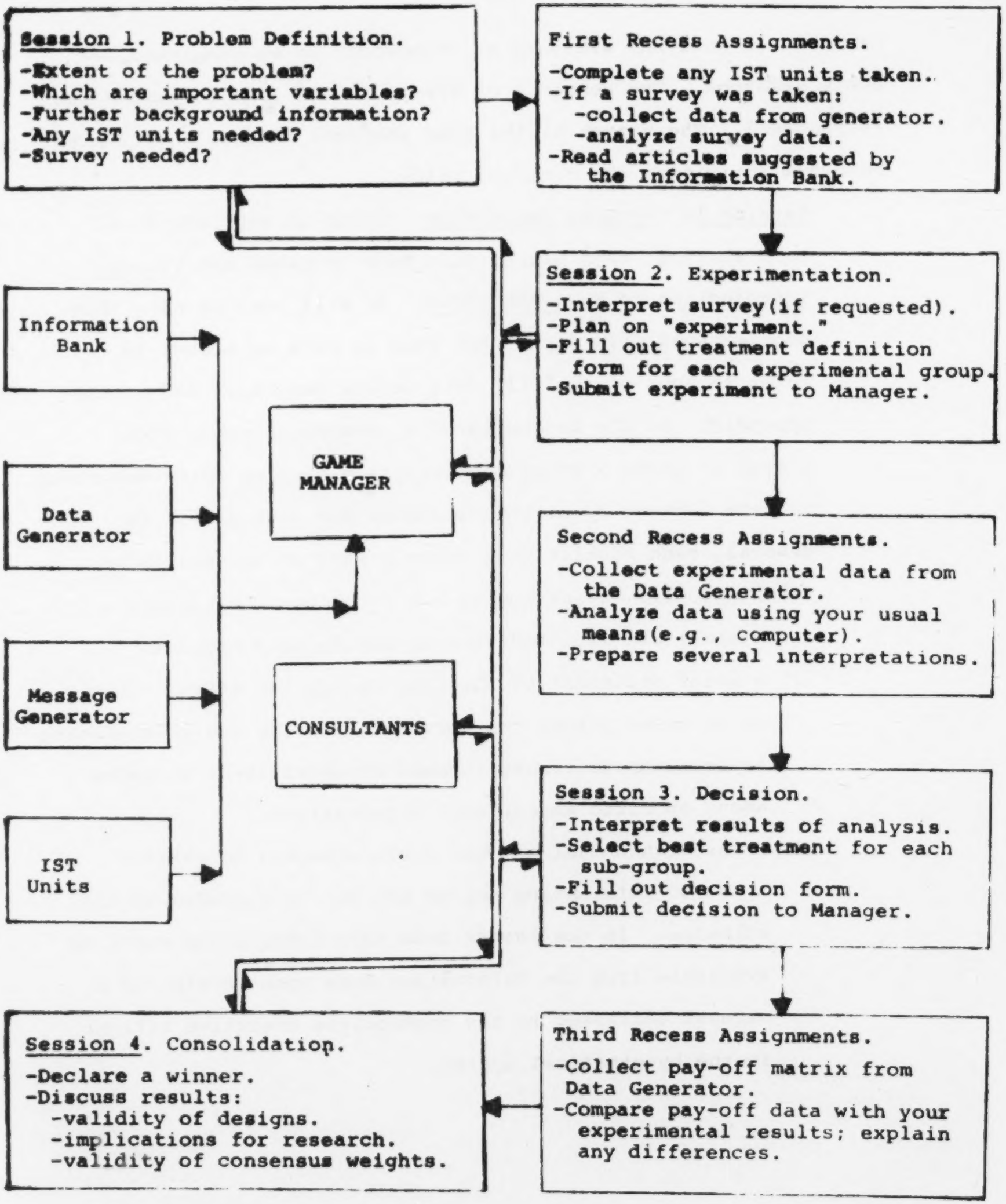


FIGURE 1. SUMMARY CHART OF FEHR-PRACTICUM GAME

statistical analyses assigned as "homework" to be completed between sessions. The sequence of events in the game is summarized in Figure 1. The nature of the four sessions and the intervening homework activities are outlined below.

Session 1: Problem Definition. Prior to any individual game session, each player will have received the Players' Introduction to FEHR-PRACTICUM. He will need to have this booklet available during the game in case he wishes to refer to some of the Fair City census data included in the appendix. At the beginning of a particular game, each player is given a Problem Description Booklet which describes in detail their specific assignment for that game. In general, each booklet will contain each of the following:

(1) A specific definition of the role that the players (teams) are expected to play during this problem.

(2) A brief statement of the task facing the teams. This would usually take the form of gathering and interpreting information necessary to make an intelligent decision among specific educational alternatives.

(3) A research budget and the costs attached to various types of information may or may not be included in the stimulus. In the latter case this information would be available from the Information Bank upon receipt of a request addressed to the appropriate executive officer in the hypothetical system.

(4) A brief summary statement of the rules of the game.

For readability this description will present the rules in context with the various game components to which they refer.

The player's task during this session is to select from the set of variables (tests) listed in the problem description booklet those which are relevant to this problem, and to specify the probable relationships among variables. In some of the simpler problems players may be able to do this on the basis of the background information supplied in the booklet. In most cases, however, they will want to get additional data. For example, they may wish to check on the validity of one or more of the "tests," or to examine the findings of previous investigators with respect to their problem. This type of material is stored, in condensed form, in the information bank.

If more comprehensive information is desired, it may be desirable to conduct a survey. In a problem involving a proposed remedial reading program, for example, it may be necessary to survey the student population in order to determine the number of students involved. The procedure for conducting such surveys is explained in detail during the orientation session.

Assignment. Those players (teams) who requested a survey during session 1 must collect the survey output from the data generator and analyze the data before the beginning of session 2. In addition, those players

(teams) who had opted to take an IST unit would normally complete it during this recess.

Session 2: Experimentation. The player's task during this session is to devise a research design which will give him valid information on which to base his decision. He will, of course, have to prepare a budget and plan his experiment within its restraints. At the end of the session he must submit an "experiment" to the data generator. Again, the exact procedure for conducting such simulated experiments is explained during the orientation session.

Assignment. Each player (team) must collect the results of his (their) experiment and analyze the data before the beginning of session 3.

Session 3: Decision. The player's task during the third session is to interpret the results of their statistical analysis of the experimental data. At the end of the session he must submit to the data generator a decision which specifies the treatment or treatment combination which he considers best. Again, the details of the method by which this is done are explained during the orientation session.

Assignment. Before beginning session 4, each player (team) must collect the "pay-off matrix" which results when his (their) decision is fed to the data generator.

Session 4: Consolidation: The FEHR-PRACTICUM game provides players with two kinds of feedback on the quality of their performance. The first of these is a "pay-off matrix" which is computed by the data simulator. When a player's

decision is entered into the data generator, it computes for each dependent variable (i.e., each variable which might be changed by a treatment) the population mean before the decision was implemented and after the decision had been operated for one academic year (40 weeks). In addition, the program computes a t test of the significance of the difference between each pair of means. This material is printed out as a pay-off matrix.

In addition to the pay-off matrix, a composite criterion score is computed for each player (team). When the problem was prepared, a team of experts were asked to rate the importance of each available variable for the decision which the players were asked to make. Using these ratings, a consensus weighting for each variable was obtained. A player's composite criterion score is computed by multiplying each dependent-variable mean in his pay-off matrix by the respective consensus weight and summing over all variables. The player or team with the highest composite criterion score is declared the winner of the game.

A second type of feedback is provided during session 4, when the game manager and research consultant(s) meet with all the players (teams) to discuss the quality of their decisions. Using the pay-off matrices as a guide, they discuss the relative merits and demerits of the research procedures used by each group. Throughout the consolidation session, the discussion focuses on the means by which each player (team) could have made a more valid decision.

SECTION II. TECHNICAL DETAILS

This section describes the technical details of the message and data generators. It includes a rationale for the particular teaching strategies adopted as well as some data supporting the validity of the underlying models.

THE MESSAGE GENERATOR

FEHR-PRACTICUM is conceived as bridging the gap between coursework and practical experience by providing the student with a computer simulation of real research without its expense and time commitment. Given the goal of simulating the research experience, common sense would dictate that the closer the simulation is to reality, the more valuable will be its contribution. The concept of the Message Generator was developed with that thought in mind. If research were a straightforward application of knowledge, one major source of researcher frustration would not exist. But in the real world, "the best laid plans....". Computers break down. Funding agencies cut budgets. Subjects in longitudinal studies disappear mysteriously. Various events in the real world, randomly delivered, do affect the course of research, demanding adaptation. The purpose of the Message Generator is to simulate this aspect of the real world.

Description

The Message Generator was originally conceived of as a set of interrupts. At each interrupt, one of a large set of stimulus events would be selected at random. Such interrupts would occur

at random times during the course of the game. Interrupts were to be short messages explaining some monetary cost or benefit, e.g., "transportation for regional students, assess $150.00," or "new budget necessitates 10 percent cut in funding." Messages could be positive as well, e.g., "computer time available through University, save $200." But the consequence of every message would be directly monetary. Any effect on the research itself would be indirect. For example, cost attachments might necessitate the use of a smaller sample, a cheaper test, or using scores already obtained instead of running a new test.

Yet, in the real world, not all random factors are directly monetary. As mentioned earlier, subjects could be lost from the sample, new literature could appear which questions a particular testing instrument, or a teacher strike could terminate an experiment prematurely. As a result of these considerations, a category system was developed for the messages. It runs as follows:

Class One - Informative messages.

These messages would consist of the most current relevant literature in the content area of the particular problem or information on tests and statistical methods. All information would be current and therefore supplemental to information in the information bank.

Class Two - Messages demanding action.

These messages consist of requests from various interested parties (e.g., superintendents, chairmen of school boards, research directors, principals,

etc.) for various actions which cost time and some money. Examples might include preparing reports, surveys, workshops, and consultations. A further refinement in this category would be to introduce response contingencies (unknown to the team). Thus, failure to produce a response to a request might produce some ill-will resulting in a subsequent penalty (e.g., a budget cut).

Class Three - Actions of others which have an effect which is not directly monetary.

These concern occurrences in the community that affect the course of the experiment in non-monetary ways. Some examples are racial problems leading to subjects being withdrawn, computer malfunctioning, or cooperation problems (perhaps contingent on messages in Class Two).

Class Four - Messages with monetary consequence.

This category includes all messages with a direct penalty or benefit in financial terms, such as budget cuts, increases in data processing costs, etc. Examples of positive messages might include donated computer time and free consultation.

Operation of the Message Generator

It was decided that one message should be used by all teams to ensure comparability of team performances. Otherwise, one team might be sorely handicapped by a message which the rest did not receive. It was also decided that individual message sets

should be developed for each problem. This allows us to develop specific situations which are more relevant to a particular problem.

The time at which messages should be delivered is still under consideration. This will be determined empirically by experimenting with various delivery schedules. The steps of the game when they could be delivered are as follows:

1. When the team requests information from the bank.
2. When the team receives information.
3. Before the team submits its research design.
4. After the team submits its research design.
5. While the team is analyzing data.
6. After data analysis is completed.
7. At the beginning and/or end of each of the first three sessions.

The feasibility of interrupts at some of the above times is dependent on the data generator's ability to produce data consistent with a given set of conditions for a number of simulated weeks, stop when the message indicates a change has occurred, and continue generating data consistent with the new conditions.

Status

A complete set of messages has been developed for the AUTOMATH problem. These include a set of general messages which can be used with any problem. In general, the messages are in the form of memos, letters, and releases.

Direction for Evaluation Year

Tasks in the future include developing specific messages for other problems, coordinating message delivery with the data

generator, and identifying proper times for delivering messages. It also remains to work out by trial and error the best format for handling contingent messages. We are currently printing the consequences of various actions on the back of master cards held by the game manager and allowing him to administer the penalties and awards.

FEHR-PRACTICUM DATA GENERATOR

The Data Generator consists of four separate subroutines within a general management program. These subroutines are termed Database, Survey, Treatment, and Decision. Each of these is discussed in turn.

I Database. We had considered it desirable to have a theoretical population (database) of some 10,000 individuals. Because of storage limitations we could have only 500 unique vectors of criteria and predictor scores. We wanted to simulate populations of over 10,000 individuals but at the same time be able to get to the unique scores of any one individual consistently. Ten thousand sets of unique scores was too much of a storage requirement for direct addressing on the computer.

We therefore decided to generate a limited number of blocks of 50 individual vectors, with each score vector having the potential of representing a large number of individuals. Since the vectors in each block were generated using the same parameters, each block is somewhat homogeneous and has the capacity of simulating common group characteristics. Therefore, groups having similar characteristics (e.g., students from a deprived environment who have a common ethnic background) are represented internally by the same block. Several groups which differ in the mean values of these characteristics but which have the same interrelationships among variables can be generated from the same block. Groups with different interrelationships among characteristics are generated from different blocks. The database subroutine generates ten blocks of 50 individual score vectors

with each block having a different pre-specified pattern of interrelationships among variables.

Paramaters needed for subroutine:

i) number of continuous predictors
ii) number of discrete predictors
iii) for each continuous predictor a correlation with two separate factors ⟨A & G⟩ and the proportion of weight to be given to ⟨A⟩
iv) the number of categories for each discrete predictor and the probability of each category
v) the number of criterion variables
vi) relative weights of each variable for the multiple regression generation of the criterion variables
vii) the proportion of variance accountable by the regression for each criterion

Algorithm. Predictors are first generated and these predictors are then used to generate the criteria. Criteria that have already been generated may be used in the subsequent generation of other criteria.

i) Generating continuous predictors

Let X = predictor to be generated as a standard normal deviate
RA = correlation with hypothetical factor A
RG = correlation with hypothetical factor G
WTRA = weighting given to factor A
A = the Z-score value of factor A (selected at random and constant for any one individual)
G = the Z-score value of factor G (selected at random and constant for any one individual)
Z RAND = a randomly selected Z-score from a uniform distribution of Z-scores according to percentile

then

$$X = WTRA * (RA * A + \sqrt{1 - RA^2} * Z\ RAND) + (1 - WTRA) * (RG * G + \sqrt{1 - RG^2} * Z^1 RAND)$$

Z and Z^1 are selected independently

ii) Generating discrete predictors

Let k = number of categories

$p(\ell)$ = probability of category ℓ such that

$$\sum_{\ell=1}^{k} p(\ell) = 1$$

randomly select a category i according to the probabilities p(i)

then the discrete predictor vector has value

$$\{\delta_{1i}, \delta_{2i}, \text{---} \delta_{ki}\}$$

where $\delta_{ji} = \begin{cases} 0 \text{ if } j \neq i \\ 1 \text{ if } j = i \end{cases}$

iii) Generating criteria

Let Y = value of the criterion to be generated (as a standard normal deviate)

w(i) = the relative weight of predictor i in generating Y (if the predictor is discrete then there is a weight for each category). (Any criterion already generated can have a relative weight and enter into the prediction of subsequent criteria)

wt(i) = normalized weight of predictor i in generating Y

$$wt(i) = \text{sign } (w(i)) \left| \frac{\sqrt{w(i)}}{\sum_i w(i)} \right|$$

Ry = the proportion of the variance of Y accountable by the regression of all the predictors

then

$$Y = Ry * (\sum [wt(i) * x(i)]) + \sqrt{1 - R_y^2} * Z \text{ RAND}$$

Notice that all continuous variables are represented internally as Z-scores. All discrete variables are represented internally as a vector of 0's with category chosen having a value of 1. One block of scores (50 individuals) is generated with the same set of parameters iii) thru vii). For each block, parameters iii) thru vii) are varied.

Status of Database. The program is written and completely operational.

Validation of Database

(i) For generating continuous predictors:

The algorithm used is a simple extension of one described by Wherry, et. al. (1965). Analysis of outputs from our pilot data indicates that the predictors are indeed standard normal deviates. For example, one generated variable had a mean of .0038, a standard deviation of 1.22, and coefficient of skewness of .077, and a standard error of .17 for 50 observations.

(ii) For generating discrete predictors:

Observed proportions of categories are consistent with the probabilities of those categories that we specify. For example, with 50 cases one generated dichotomous variable had observed proportions of .7 and .3 for our specified probabilities of .65 and .35. Another variable with 4 categories had proportions of .4, .34, .10, and .16 for our specified probabilities of .35, .35, .15, and .15.

(iii) For generating criteria:

The algorithm used is an extension of the work of Bottenberg and Ward (1963) in combination with the one of Wherry, et. al. (1965). Analysis of outputs from our pilot data indicates that the criteria are standard normal deviates. For example, one generated criterion had a mean of .38 (standard error of mean = .13), a standard deviation of .95, a coefficient of skewness of -0.023 (standard error = .34), and a coefficient kurtosis of -0.36. Another had a mean of .54 (standard error of mean = .16), a standard deviation of 1.10, a coefficient of skewness of -0.17 (standard error = 0.34), and a coefficient of kurtosis of -0.83. These values are all well within the range expected from sampling variations.

We ran multivariate regressions on our pilot data using our predictors (both continuous and discrete) as independent variables and each of our criteria in turn as dependent variables. In four instances where we specified values of the multiple R equal to .70, .90, .65, and .85, we obtained values of .68, .92, .77, and .88, respectively. Again, these values are well within expected sampling error.

The regression coefficients for each of the independent variables bore a strong resemblance to the nominal weights on the predictors that we specified for generating the criteria. This relationship was not intended to be linear because we desired to simulate the effects of (unknown) moderator variables. We can, however, manipulate the relative importance of each predictor using the relationship between nominal and effective weights given by Wang and Stanley (1970).

In our pilot data almost all coefficients for predictors had the same sign as the nominal weight. The exceptions were predictors with very small relative weights--a situation where some random error is sure to be present to influence a possible change in sign. To test the validity of our prediction algorithm further we obtained correlation matrices for our pilot data. The correlations between predictors and criteria agreed in sign with the signs of our specified weights. We then changed a single weight and observed the resultant change in the correlation matrix. The result supported the validity of our algorithm. For example, when a relative weight for a predictor was changed from +10 to +99 while all other weights remained constant, the primary change was a large increase in the correlation between the changed predictor and the criterion (see Table 1). This was considered convincing evidence that we controlled the strength of predictive relations.

TABLE 1. EFFECTS OF CHANGING NOMINAL WEIGHTS

Predictor	Wt. for Predictor in Generating Criterion	Correlation Between Criterion & Predictor	Wt. for Predictor in Generating Criterion	Correlation Between Criterion & Predictor
1	+10	.016	+99	.265
2	+05	.025	+05	.137
3	+90	.296	+90	.249
4	-17	-.259	-17	-.109
5	-05	-.163	-05	-.203
6	-08	-.315	-08	-.303
7	+15	.142	+15	.123
8	-15	-.033	-15	-.076
9	+05	.158	+05	.177
10	+10	-.221	+10	-.236

Rationale for the Generation of Rawscore Matrices

A major objective for the program was the capacity to treat the same individuals (i.e., change their score vectors) on successive occasions. It was, therefore, necessary that any one individual be represented by a vector of scores within one block and that any reference to this individual always pointed us to this same vector. Each individual was given an identification number consisting of a two digit group reference number, a two digit sub-group reference number, and a three digit individual reference number. In some problems it is appropriate to think of the group as being a school and the sub-group as being a class within the school. So identification 10-14-017 refers to individual 17 in class 14 of school 10.

For each class of each school we store a pointer to one of the blocks of 50 scores that have been generated. Were we simply to use the line number of the referenced individual in obtaining his scores (e.g., the 17th vector in the block for individual 17) strong patterns could begin to emerge that would destroy the validity of the simulation. So we attempted to 'scramble' the one-to-one correspondence of individual identifications with order of score vectors. We do this by providing a key number for each class that is used in the score vector finding algorithm. This key number is always a 3 digit prime number. When this key number is multiplied by the individual identification number and the result divided by 50, the remainder of the division is a number less than 50 and can be used to reference one of the score vectors within a block. We use prime numbers because they produce a string of 50 remainders before any repetition occurs.

Each internal Z-score of the score vector can be used to refer to more than one real-world or external score. For instance, an internal Z-score can stand for verbal intelligence. This internal score can be used to generate a number of external scores such as the language score on the California test of Mental Maturity, the information subtest on the W.I.S.C., the comprehension subtest of the W.P.P.S.I. or the sentence completion subtest on the Lorge-Thorndike. We handle this in the computer program by having each external variable indexed to one internal variable. Of course more than one external variable can be indexed to an internal variable. For each external variable we provide a mean (μ), a standard deviation (σ) and a reliability measure (r). Assume that the internal variable to which the external is indexed has a Z-score of Y. Then the external value is

$$X = \sigma[rY + (\sqrt{1 - r^2})\ Z\ RAND] + \mu$$

We specify a number of vectors of means--each component of the vector being the mean value for one external variable. We do the same for standard deviations and reliabilities. By having more than one vector of each we can have different vectors refer to different subgroups. So the mean value of a grade 1 class in a certain school on a certain external variable can be made different from the mean value on that same variable for another class.

The effect of all of this is to enable us to simulate and generate a very large database without having to store it all in computer memory. Instead we generate the blocks of lines of internal Z-scores and use one line of the Z-scores to characterize

an individual. We have a number of vectors of external means, standard deviations and reliabilities which can be used in various combinations to produce actual external scores. And by this scheme we are able to always generate the same external score when requested for any one individual, or a similar score if the reliability of the measurement is less than 1.

The capacities discussed above are provided by the Survey and Treatment subroutines. Each of these in turn is described below.

II Surveys and File Searches. We enable experimenters to sample from the data bank which has just been described with the survey subroutine. We can specify a set of scores which exist in the school files and which can be obtained at a specified low cost. Others in the set of scores do not exist but must be measured by the "administration" of tests as specified by the experimenter. (He can also choose to remeasure certain of the variables that exist in the files.) Of course, each test costs something. If a request for information from the files on a set of individuals is made we call that procedure a file search. If a request is made to measure some specified set of variables for individuals, then we call that procedure a survey.

For any file search or survey we provide a clerical service to the experimenter. For a certain fixed cost the computer will sort through all the scores requested for all the individuals in the search or survey and print out only the scores for individuals that meet the set of conditions specified by the experimenter. For example, if the experimenter asks that only the scores for

high socio-economic white males with WISC's less than 90 or low socio-economic black females with Stanford-Binets greater than 130 be printed from his survey he can obtain that service from the computer simulation at a small fixed cost. This can save the experimenter much effort if he plans ahead and makes use of this service.

Status of Survey

Algorithms have been written and tested. We are currently integrating these algorithms into the master program.

Validation: We have tested the algorithms for the clerical service and they perform as required: only the scores of those individuals which satisfied our test relations were printed out.

III Treatments. Experimenters can request that treatments be administered to individuals. The possible treatments are given in the descriptive materials for the game. The experimenter selects the sequence of treatments he would like and the duration of each treatment. He does this for each treatment group he sets up. Simultaneously, he specifies which external variables he wants measured and at what times he wants them measured.

The computer simulates the effects of the treatments, which are affected by the states of each of the internal variables. The state of each internal variable enters into the computation of a time modifier which is used to change the experimenter's specified time into a base unit for the growth curve. Weights are specified for each internal variable with each treatment so that the values of some variables are more important than others in determining the effects of the treatment.

For example, let:

v(i) be the value of variable i

w(i,j) be the weighting of the value of variable i in affecting the outcome of treatment j

N be the number of variables

Then the time modifier (TM) is:

$$TM = \frac{1}{N} \sum_{i=1}^{N} \left[\frac{v(i) + w(i,j)}{w(i,j)} \right]$$

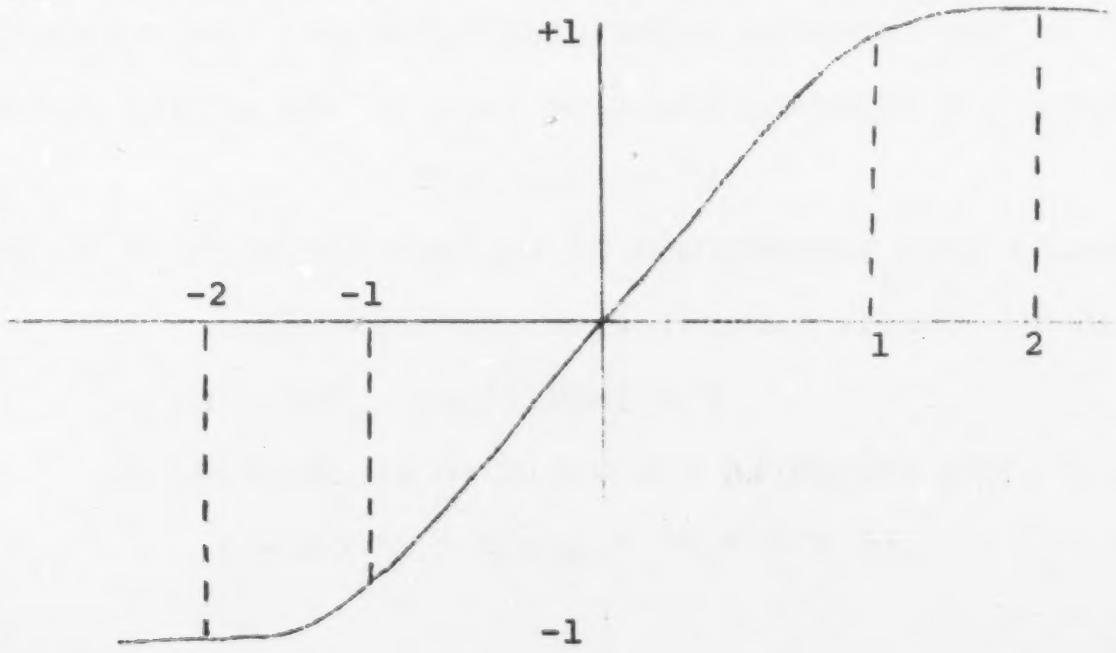
The base unit used in the growth curve is:

T = TM * TIME

where TIME is the experimenter's specified time [i.e., the actual time the treatment is to be applied.]

As mentioned, each criterion variable has a certain 'growth curve' when a treatment is administered for a given duration. This growth curve is obtained from shifts in the origin, rotation of axes, and modification of the scales of the axes of the hyperbolic tangent. The growth curve gives us the algebraic change in the variable we wish the treatment to modify. In all instances a treatment time = 0 will lead to no change at all.

Consider the curve of the hyperbolic tangent y = tanh x



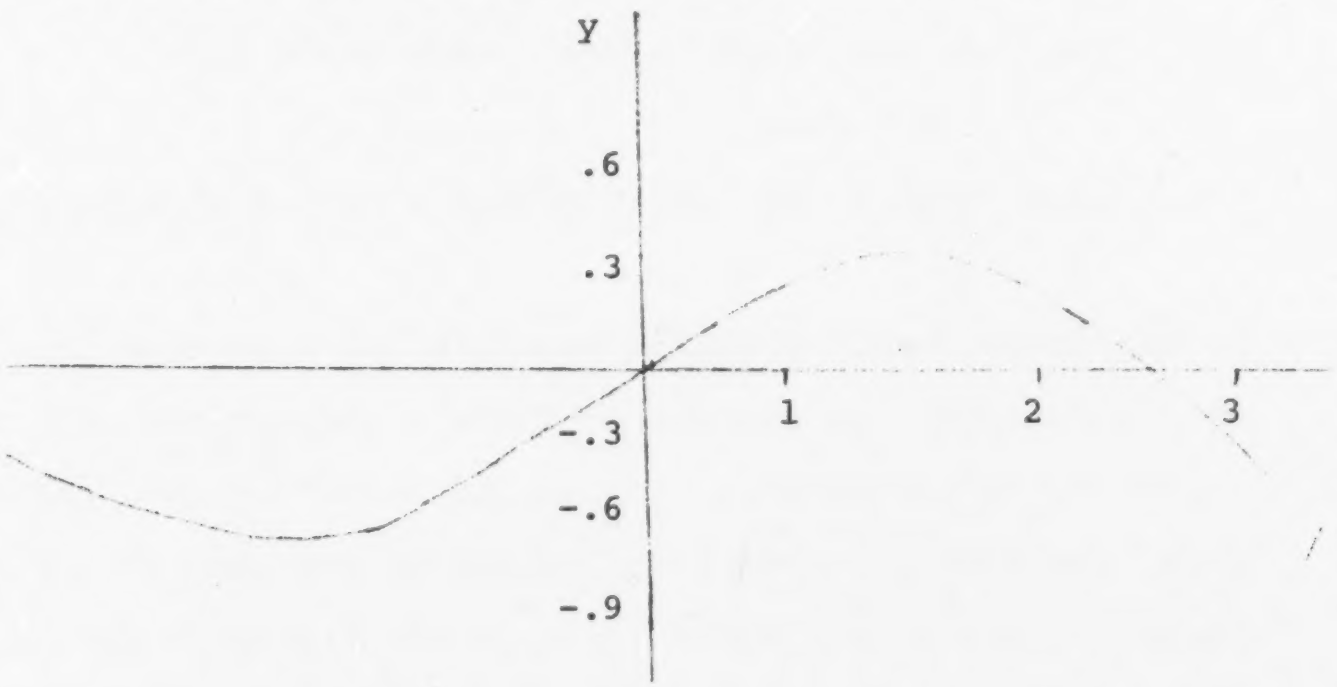
We can move the origin of the curve by specifying that it be at (-a, tanh -a).

We can rotate the curve through an angle of θ radians clockwise.

We can shrink or expand the time scale by multiplying it by b.

We can shrink or expand the y-axis by multiplying it by c.

Then one possible modified growth curve looks like:



On this modified growth curve a time T = 1 will lead to a change of .3 in the variable under consideration. We accomplish this by first transforming the time unit on the growth curve.

$$T^1 = T*\cos \theta/b$$

And using this transformed T^1 to find the value Y on the untransformed hyperbolic tangent with shifted origin

$$Y = TANH(T^1-a) - TANH(-a)$$

Finally, the change in the variable is obtained as

$$YC = C * T^1 * \sin \theta + Y * \cos \theta$$

This procedure enables us to simulate almost any growth curve and to maintain a unique growth curve for each subject on each criterion variable. The procedure is approximate but quite appropriate for our purposes. We use it because it is computationally convenient. Since we do not scrupulously adhere to the notion that all growth curves are modifications of the hyperbolic tangent we have no qualms about making approximations where it is convenient to do so.

We simulate the effects of testing by changing the effective lengths of the treatments. For example, a treatment duration of 4 months will lead to a different result than a treatment duration of 2 months followed by a test followed by another 2 months of treatment.

The carryover effects from one treatment to another were more difficult to handle. There is at present no database from which to derive their nature. However, it seems reasonable that treatments applied in sequence have no direct effect on each other. Any differential in the effects of a treatment given second in a series rather than first can probably be explained by the changes in the entering characteristics of individuals who have experienced a prior treatment. In our program, each treatment applied makes some changes in the vectors of variable scores which represent the subjects. Since treatment effects are interactive with the initial values of these vectors, a treatment applied second in a sequence will result in different outcomes than if it were applied first. In this way, the program introduces a somewhat unpredictable carryover effect.

Status: The treatment program is written and is operational.

Validation: Algorithms here are quite straight-forward. In trial runs we have been able to obtain the same set of scores for the same individual on successive runs. Patterns of scores are different for different treatment groups, and the direction and magnitude of group differences are controlled by our treatment parameters.

The validation of the algorithm for unique growth curves for individuals is not yet complete. However, the preliminary results indicate that the individual's raw scores fit his specified curve within the limits of the "measurement error" which we specify as a test reliability parameter. We have been able to use fictitious data and have a reasonable set of results simulated. By reasonable we mean that the scores do demonstrate growths as a function of time in accordance with the growth curve that we specify. These growths are within specified bounds too. We can also simulate carryover effects, but because of the novelty of this approach, we do not know yet what weights should be assigned to the constellation of states of the variables to give us a realistic set of results. So, while we can presently specify fictitious growth curves that we can use in the program we are currently striving to make these treatment effects consonant with the real world. Unfortunately, there is very li'tle in the educational literature about the influence of a set of individual variables on the effects of treatments to help guide us in this approach.

IV Decision. The final phase of the game consists of having the players make recommendations about treatment programs to be applied to the groups they designate. They are able to specify up to four different treatments. A fifth treatment--the control--is automatically provided for all individuals not assigned to any of the previous four. Control in this sense means the treatment which is presently being used.

When making recommendations about treatment programs players indicate the characteristics of the individuals who will be assigned to each program. They are not allowed to make the recommendation in terms of individual I.D.'s but must describe the entire subgroups in terms of variable scores; for example, individuals with a WISC IQ less than 80, and a language score on the California Test of Mental Maturity less than 100. This requires the players to go beyond the immediate data and to think of the variables that are of importance to the success of each treatment program.

In calculating the effects of a decision, we use the same algorithm to sort individuals into treatment groups as was used to sort individuals in the clerical service option for the file search. The main feature here is that every individual in our simulated population must be assigned to one of the treatments. The specified sequence of treatments is then administered to every individual in each group, using the treatment algorithms to generate the score changes in the same manner as described previously, but with the treatment time set by the control parameters we provide rather than by the players. Here there is only one time period for testing--at the end of the treatment period. Thus

test results are comparable over all possible decisions. In addition, a constant measurement time circumvents much debate over the proper time to measure the effects of a treatment. This is necessary in order to provide a determinate game winner. We plan to experiment with various evaluation times (1 year, 2 years, etc.) during the second year of the project.

The "payoff matrix" provides to each team a set of summary statistics on the results of the implementation of their recommendations for the set period. For each treatment group we print the group size (N), and compute prior and posterior means and standard deviations for the pre-determined set of variables which we have defined internally as experimentally important. In addition, we provide the matrix of intercorrelations among these scores. We also compute overall means and standard deviations (prior and posterior). We provide all this information to the game manager who uses it during the consolidation period.

Status: The decision subroutine is now completely operational.

Validation: The decision subroutine is a straight-forward application of the sampling and treatment routines to the entire population. To the extent that the preceding algorithms are valid and operational there is no problem of validity for this phase. The realism of the payoff matrix is not yet evident. We hope to involve a number of cognate specialists early in the second year to help us determine whether the system as presently used provides valid information about the quality of a particular decision.

Portability of the Computer Program

The computer programs were written in the American National Standard (ANS) version of FORTRAN IV to optimize their portability from one computer installation to another. However, different computer manufacturers do not support identical system routines within the standardized FORTRAN language. For example, the methods by which a particular record can be retrieved from a large file varies from system to system. To minimize problems of this type, we restricted the instructions used in the FEHR-PRACTICUM data generator to the subset of instructions which are common to most compilers. As an aid in this task we used Appendix B of Golde (1966) which lists the characteristics of FORTRAN IV compilers for a wide variety of computer models.

A second precaution introduced in the interests of portability was to minimize the dependence on live memory. Instead, we store the massive matrices used to store our research population on various tape drive units (or disks, if applicable). Consequently, we believe that the program should be operable on the computer systems available to most medium-sized colleges. It remains for the proposed second year of this project to validate this claim.

DEVELOPMENT OF PROBLEMS

The original proposal committed us to the development of ten problems in a variety of educational settings. The process of problem development was divided into four stages. The following section will briefly describe each of the stages.

I. Selection of Problem Areas: Criteria

The problem topics were selected on the basis of certain criteria which were believed to make the game both instructive and interesting to the players.

Problem areas were chosen according to the following guidelines:

a) The problem was common in the real world.

b) Data on relationships among variables, and some research to validate that information were available. This information could then be programmed into the computer.

c) Experts were available to provide the detailed descriptions which were needed in the problems. In particular, we needed details respecting costs, personnel needs, and the type of problem likely to arise during research.

d) The set of problems should represent a wide range of interests in educational research. We wished to ensure the players a good cross-section of the problems encountered in the real world. Ten problems were chosen to be developed.

II. Problem Description Packages

After the problems had been selected, all information which would be needed by the players was collected. For each problem,

this consisted of a description of each of the following:

a) The setting in which the problem took place.
b) An itemized list of the costs attached to a problem, and specific information on the players' role in the problem.
c) A "catalog" listing the "tests" which could be used in the solving of the problem. In addition, a detailed description of each test that was prepared and stored in the Information Bank.

III. Developmental Process

Typically, problem development was carried out in two phases. In the first phase, a wide search of the literature was made. From the literature a list of important variables was compiled. Then five experts in each of the problem areas were asked to add any variables they felt were important. We also asked them to indicate irrelevant variables on our compiled list. This task will be completed by August 30, 1971. From the above process two separate variable lists were compiled: the common variable list, which contained those variables common to all ten problems, and ten specific variable lists (one for each problem) containing variables unique to the individual problems.

IV. Data for the Information Bank

To optimize the utility of the game much of the reference material necessary to devise an appropriate research design is stored in the information bank. Consequently, problem development includes the preparation of the relevant index cards for the bank. These include the following:

1. Summaries of the most relevant research information in the problem area itself.
2. For each test available, a listing of the purpose of the test, the age-range for which the test is applicable, time for administration, cost, number and kind of subtests (if any), and the types of scores obtained.
3. Statistical information on all tasks and subtests (e.g., means, standard deviations, reliabilities, and validities for all age (grade) levels).
4. Summaries of critical reviews of each test.
5. A listing of the name of each school used in the problem; its size, number of teachers, number of classrooms, etc.

V. Internal Representation in the Computer

Once the variables have been chosen, we must specify how each variable interacts with every other variable. In the AUTOMATH problem, for example, we need to know what previous research has shown respecting the interactions among social class, classroom placement, sex, and the like. The task is not technically difficult, but it is long and tedious. These data are then used to determine the values to be entered as parameters in the treatment program (described in the previous section). The literature search is now completed. The translation of the findings into population specifications (parameters for the program) was a task specifically reserved by the original proposal.

TRIAL OPERATION

A trial operation of the total game was undertaken in early July, using the members of the project staff as subjects. The relationships defined in the AUTOMATH problem were translated to

parameters, and a complete game played using Mr. Smith's requests as outlined in the Orientation booklets. We were able to obtain results which were consistent with the theoretical relationships reported in the literature only after several trial-and-error readjustments of the parameters.

Originally, we had planned a subsequent run of the game using subjects unfamiliar with the project. However, the trial run using the staff as subjects revealed an important weakness in the program which should be corrected before proceeding with a trial run involving "real" subjects.

The previous method of distinguishing between information held in a file, which could not be changed by a treatment or by measurement error, and the same test when it was readministered during an experiment was to represent them by different internal variables. We found that this procedure was both wasteful of storage space and far too cumbersome to program. Consequently, it was decided to revise the program to represent both the unchanging file values and the variable scores taken during an experiment by the same internal variable. We accomplish this by a simple sub-program which ensures that any external reference which is made during a file request will always generate the error portion of a score from the same random number sequence, while a request to that variable which is made during a survey or treatment request has an error term generated from a variable random number sequence. Thus, a file request will always yield the same score for a particular individual, while the survey or treatment will yield scores which differ within the limits of the standard error of measurement for the variable concerned.

The rewriting task is now completed, and a trial run of the AUTOMATH sample problem using subjects unfamiliar with the game is now scheduled for the first week in August. We plan to issue a supplement to this report when that trial operation is completed.

SUMMARY

The purpose of this project was to develop the physical components of the proposed FEHR-PRACTICUM game, and to demonstrate that the game, as a whole, was operable. Evidence was presented that each of the components, individually, operated according to specifications. A trial run of the game, conducted in July, demonstrated that the game model was operable, and suggested several modifications for smoother operation. These modifications are now completed, but further testing of the model depends on the funding of a second project to complete the second year of the two-year project outlined in the initial FEHR-PRACTICUM proposal.

Although much remains to be done before we can claim that the game produces a valid representation of the real world, the mere existence of an operational game model requires some mention of the possible future developments of that model. A brief discussion of three of the possible extensions of the FEHR-PRACTICUM game are outlined below.

Training Educational Decision-Makers

A case can be made for the proposition that the effective use of the findings of educational research requires educational decision-makers who are trained in the effective use of the research facilities at their disposal. We suggest a module which trains administrators and researchers as a team, with the FEHR-PRACTICUM

game simulating the environment for both. This could be accomplished by expanding the message generator to include a large number of in-basket items, with a relatively small proportion of the items requiring research activity.

Studying Problem Solving

If the game proves to be a valid representation of real-world research, the present model should provide an ideal vehicle for studies interested in comparing the characteristics and research activities of players who arrive at "good" decisions versus those who arrive at bad decisions.

Behavior Modification

It should be possible to create a set of variables which represent occurrences (or non-occurrences) of particular events and to functionally relate these to a large number of teacher behaviors (treatments) which, again, either occur or do not occur. Students could interact with the model, choosing a particular teacher behavior on the basis of some response from his (or her) class.

The general idea behind this notion is being tried out in the pre-school problem to be validated for the second year. If this proves to be successful, it should be a relatively simple matter to modify the game to the fully interactive system suggested above.

REFERENCES

Bottenberg, R.A. and Ward, J.H., Jr. Applied Multiple Regression. Clearinghouse for Federal Scientific and Technical Information, Technical Documentary Report PRL-TDR-63-6, 1963.

Buros, O. (editor). The Sixth Measurements Yearbook. Highland Park, New Jersey, 1965.

Buswell, G.T. and McConnell, T.R. Training for Educational Research. Center for the Study of Higher Education, Berkeley, California, 1966.

Collet, L.S. SIMEX: A Program to Simulate Experiments for Teaching and Testing Research Competencies. Association for Educational Data Systems Journal 1(3): 10-18, September 1969.

Golde, Hellmut. FORTRAN IV for Engineers and Scientists. Macmillan Company, New York, 1966.

Harshbarger, Thad R. Introductory Statistics: A Decision Map.

Harshbarger, Thad R. Workbook to Accompany Introductory Statistics: A Decision Map. Macmillan Company, New York, 1971.

Wang, M.W. and Stanley, J.C. "Differential Weighting: A Review of the Methods and Empirical Studies." Review of Educational Research 40(5): 663-705, December 1970.

Wherry, R.J., Sr.; Naylor, J.C.; Wherry, R.J., Jr.; and Fallis, R.F. "Generating Multiple Samples of Multivariate Data with Arbitrary Population Parameters." Psychometrika 30(3), 1965.